ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 39,No. 3,May2024,pp. 761－774
DOI:10. 16337/j. 1004-9037. 2024. 03. 022

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

移动对象室内定位中的隐私保护方案

张志武，雷若兰，乐燕芬

(上海理工大学光电信息与计算机工程学院,上海 200093)

摘　要：针对Paillier算法应用于室内指纹定位隐私保护时,计算开销大而影响定位实时性的问题,本文提出了一种移动对象室内指纹定位中的隐私保护算法,解决Paillier加密算法实时性不足的问题,同时实现移动对象的轨迹匿名并有效提高定位性能。考虑到参与定位的信号接入点(Access point,AP)与参考点(Reference point,RP)的数量是影响加密算法时间开销的主要因素,算法将对象的轨迹定位分成连续位置定位与非连续位置定位。连续位置定位利用前后定位请求的信息减少参与加密运算的AP与RP数量,而非连续位置定位用户利用粗定位减少算法涉及的AP与RP数量,进而提高定位效率;并提供一种主成分分析法(Principal component analysis,PCA)的可选方案,进一步提高定位效率。实际环境中的定位实验结果表明,所提算法在连续位置定位与非连续位置定位中,用户端单次定位所需时间均可控制在1 s,且连续定位中定位精度提高了约20%,非连续位置定位中采取的隐私保护措施对定位精度基本无影响。整个定位过程实现了移动对象的轨迹保护,有效提高了定位算法的整体性能。

关键词：室内定位;WiFi指纹;Paillier算法;轨迹匿名

中图分类号：TP391　　**文献标志码**:A

Privacy Preserving Scheme for Indoor Positioning of Mobile Users

ZHANG Zhiwu, LEI Ruolan, LE Yanfen

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Aiming at the problem of high computational overhead affecting the real-time localisation when paillier′s algorithm is applied to indoor fingerprint privacy protection, this paper proposes a privacy-preserving algorithm for indoor fingerprinting positioning of mobile users to achieve trajectory anonymity and effectively improve the positioning performance. Since that the number of access points (APs) and reference points (RPs) involved in localization is the main factor affecting the time overhead of the encryption, the proposed algorithm divides the trajectory localization into continuous and discontinuous location localization. The number of APs and RPs involved in encryption is reduced by using the information of the before and after requests in continuous location localization, while the number of APs and RPs involved in encryption is reduced in discontinuous location localization. In continuous position localization, the number of APs and RPs involved in the encryption operation is reduced by using the information of before and after location requests; while in discontinuous positing localization, the coarse localization of users reduces the number of APs and RPs involved in the algorithm, thus improving the

基金项目:国家自然科学基金(51705324)。
收稿日期:2023-07-07;**修订日期**:2024-03-07

location efficiency. An optional scheme based on principal component analysis (PCA) is proposed to further improve the localization efficiency. Experimental results in a real-world environment show that the proposed algorithm can control the time required for a single positioning in both continuous and discontinuous positioning within 1 s. The positioning accuracy is improved by about 20% in continuous positioning, while the privacy protection has no effect on the positioning accuracy in discontinuous positioning. The overall performance of the localization algorithm is effectively improved.

Key words: indoor positioning; WiFi fingerprint; Paillier algorithm; trajectory anonymity

引　言

随着物联网技术的兴起,位置服务成为人们生活中不可或缺的部分。而室内定位因为支持着室内环境下的众多使用场景(室内导航、室内监控等)而具有越来越高的商业价值,并受到学术界的广泛关注[1-2]。为实现更有效的室内定位,国内外学者提出各种不同的技术以在不同环境下实现对室内用户位置的精准获取[3-4]。目前广泛部署在建筑物内的无线局域网络为室内定位提供了一种实施方便、价格低廉的方案,使得基于WiFi技术的室内定位成为研究热点[5-6]。基于WiFi的室内定位通常分成离线阶段和在线阶段两个部分。离线阶段服务提供商在每个参考点(Reference point,RP)上测量来自各接入点(Access point,AP)接收的无线信号强度(Received signal strength, RSS)形成WiFi指纹,并结合相应位置记录在指纹数据库中。在线阶段用户实时测量WiFi信号强度,获得该位置的RSS指纹。定位服务提供商通过计算该位置的RSS指纹与离线数据库中每条指纹的信号相似度,获取用户的估计位置[7]。同时,随着室内基于位置的各类服务普及,人们在获取高精度的室内定位服务时,对隐私保护有了更高的要求[8]。常用的定位技术在未考虑数据的安全隐私时,请求位置服务的用户需要与服务器建立通信[9],用户需要将用于定位的信息上传至服务器,服务器完成相关定位算法后返回估计位置,在此过程中用户的信息可能遭受攻击并泄露。服务器则会受到恶意用户攻击,从而可能泄漏服务器的数据库[10]。目前,已有很多方法用于实现室内定位中的隐私保护,例如K匿名[11-12]、混合区域[13]、加密运算[14]等方法。文献[12]首次在位置隐私保护问题中尝试使用K-匿名技术,使得服务器无法定位到目标用户的确切信息,进而保护目标用户的真实位置,但移动对象在连续定位中不可避免地导致移动用户轨迹的泄漏。文献[14]首次提出基于合数阶剩余类的Paillier加密算法,除了基本的加密操作外,Paillier还可实现密文域的加法运算[14-15]。Li等[16]在室内定位中的隐私保护领域进行了开创性的工作,将Paillier加密算法应用于室内定位中,利用Paillier算法的加法同态和数乘同态实现了用户端与服务器端的加密隔离,在室内定位中实现了隐私保护,但Paillier加密算法的大数据特性使得用户端通常需要数秒才能完成单次定位涉及的加解密运算,在指纹库规模较大时,这个计算开销甚至会达到十几秒,导致定位效率低,无法保证移动对象的定位实时性。文献[17-18]中,添加了模糊逻辑技术作为隐私保护技术的补充,更多的用户导致更多的计算操作和更多的信息被传递导致现有的加密方法中的计算开销与通信开销更高。Zhang等[19]将隐私保护定位问题形式化为最小化超定线性公式的最小平方误差,然后使用超定线性表达式的特殊结构在密文空间中设计轻量级解决方案,单次定位中用户端所需时间仍需2 s左右,达不到实时定位的效果。

文献[16]将室内定位中的在线阶段分成用户RSS信号加密、服务器距离计算和用户端解密定位3个阶段,用户在加密阶段将需上传至服务器的定位RSS信息进行加密,服务器在密文域进行指纹间欧氏距离的计算并返回,用户接收服务器返回的加密距离解密后找到最近邻完成定位。通过分析此过程中Paillier算法涉及的模乘和模幂运算的时间开销,可知用户端进行加解密运算和服务器端距离计算的

时间开销均与参与运算的RP和AP个数直接相关。针对这一问题,本文所提算法通过分析移动对象在线RSS指纹的特点,判断其是否为连续位置定位,采取不同的措施减少参与定位的有效RP与AP个数,在保证定位精度的前提下,实现移动对象轨迹匿名和定位效率的大幅提升。

1 室内定位和隐私保护的相关工作

1.1 基于WiFi指纹的室内定位

基于WiFi指纹的K近邻算法(K nearest neighbor, KNN)定位为最常用的室内定位方法之一,分为离线和在线两个阶段。离线阶段,服务商在每个RP位置上采集各AP的WiFi信号的强度,构成离线指纹库。指纹库中每一条指纹为

$$\boldsymbol{\varphi}_i=[P_i,F_i]=[x_i,y_i,f_{i,1},f_{i,2},\cdots,f_{i,M}]\quad \forall i\in[1,2,\cdots,N] \tag{1}$$

式中:$\boldsymbol{P}_i=[x_i,y_i]$为第$i$个RP的物理位置,$\boldsymbol{F}_i=[f_{i,1},f_{i,2},\cdots,f_{i,M}]$为第$i$个RP的RSS信号向量,$f_{i,j}$为在$RP_i$上接收到来自第$j$个AP的RSS信号强度,通常表示采样时间内的均值,未采集到的AP信号则其$f_{i,j}$设为0或−95 dBm,N为RP的总数,M为整个定位区域内能采集到的AP个数。

在线阶段,假设用户采集到的RSS信号向量表示为

$$\boldsymbol{F}_t=[f_1,f_2,\cdots,f_M] \tag{2}$$

同时定义覆盖向量如式(3)所示,各RP或用户能接收到AP_i的RSS信号时,c_i等于1,反之等于0。

$$\boldsymbol{C}=[c_1,c_2,\cdots,c_M]\quad c_i=\begin{cases}0 & f_i=0\\ 1 & f_i\neq 0\end{cases} \tag{3}$$

服务器计算用户采集到的RSS信号与每个RP相应的RSS信号的欧氏距离,计算公式为

$$d_{i,t}=\|\boldsymbol{F}_i-\boldsymbol{F}_t\|^2=\sum_{j=1}^{M}(f_{i,j}-f_j)^2=\sum_{j=1}^{M}f_{i,j}^2+\sum_{j=1}^{M}(-2f_{i,j}\cdot f_j)+\sum_{j=1}^{M}f_j^2 \tag{4}$$

式中:$d_{i,t}$为RP_i的RSS信号与用户的RSS信号的信号空间距离,"$\|\cdot\|$"表示欧氏距离。

由式(4)计算得到当前定位请求与每个RP的欧式距离,选择距离最小的K个RP,取该K个RP物理位置的平均值作为用户的估计位置,具体可表示为

$$(\hat{x},\hat{y})=\frac{1}{K}\sum_{i=1}^{K}(x_i,y_i) \tag{5}$$

式中:$(\hat{x},\hat{y})$表示用户的估计位置,(x_i,y_i)表示选择的K个RP的实际位置,在本文实验中K取值为3。

1.2 室内定位中的主成分分析算法

在大监控区域内可能布置较多的AP,常用的降低定位算法计算复杂度的方法是按某种策略选择最优的AP集合[20]。不同于以上方法,主成分分析(Principal component analysis,PCA)算法经过线性变换从来自各AP的RSS信号中提取包含原始信息的少数几个综合指标,这些指标互不相关,也称为主成分,在降维的同时保持了信号变量的总方差不变。文献[20]将RSS指纹转换为主成分,有效地利用了所有AP的信息;文献[21]将定位区域划分为多个局部区域,并按分布密度将参考点划分为粗网格和细网格,通过各局部区域对应的信号覆盖向量和主成分分析法提取的稀疏指纹的特征实现层次化匹配。这两种定位技术均使用了PCA算法对原始离线指纹库进行预处理,以此提高定位精度和降低在线计算复杂度。PCA变换为

$$\boldsymbol{F}_{\mathrm{P}}=\boldsymbol{F}\cdot\boldsymbol{T} \tag{6}$$

式中:$\boldsymbol{F}$为变换前包含M个AP的原指纹的RSS信号矩阵,$\boldsymbol{T}$为$M\times L$维的变换矩阵,表明了每个AP对主成分的贡献量,$\boldsymbol{F}_{\mathrm{P}}$为变换后的主成分。从式(6)中可看出,PCA在降维时并没丢弃原RSS数据,而是

融合了所有AP的信息。经过PCA变换后,指纹库和用户的RSS信号均从M维降至L维。

1.3 室内定位中的Paillier算法

上述定位算法中并未考虑到室内定位中的隐私保护问题,即服务器端并不希望暴露自己采集的WiFi指纹数据库,而用户端则需隐匿自己的实际位置。Paillier算法具有的同态特性,可利用密文域的距离计算完成最近邻的检索,从而在实现KNN定位的同时达到定位中的隐私保护。

1.3.1 Paillier算法

Paillier算法分为3个部分[14]:生成密钥、数据加密和加密数据的解密。

(1) 生成密钥

Paillier算法的密钥分成公钥和私钥,公钥用于对数据进行加密,私钥用于对加密数据进行解密。Paillier算法首先生成两个大的素数p,q,公钥由n,g两个参数构成,其中n为大素数p、q的乘积,g为0~n^2的任意整数,具体可表示为

$$\text{Publickey}=\{n,g\}\quad n=pq;g\in Z_{n^2}^{*} \tag{7}$$

式中Z^{*}表示整数。

私钥由λ、μ两个参数构成,λ为($p-1$)和($q-1$)的最小公倍数,具体可表示为

$$\begin{cases}\text{Privatekey}=\{\lambda,\mu\}\\ \lambda=\text{Lcm}(p-1,q-1)\\ \mu=L(g^{\lambda}\bmod(n^2))^{-1}\bmod n\end{cases} \tag{8}$$

式中:Lcm表示最小公倍数,$L(x)$函数可表示为:$L(x)=(x-1)/n$。

(2) 数据加密

假设m为待加密的明文数据,m通过公钥进行加密后得到密文c,具体加密规则可表示为

$$c=[m]=g^{m}s^{n}\bmod(n^2)\quad s\in Z_n^{*} \tag{9}$$

式中"[ ]"符号表示加密过程。

(3) 数据解密

利用私钥对密文c进行解密得到明文m,具体解密规则可表示为

$$m=L(c^{\lambda}\bmod(n^2))\times\mu\bmod n \tag{10}$$

Paillier算法是一种加法同态,具有如下性质[14]:两个明文m_1、m_2相加后加密等于其密文c_1、c_2相乘;常数t与明文m相乘后加密等于明文m加密后的t次幂,具体如式(11,12)所示。

$$[m_1+m_2\bmod n]=[m_1][m_2]\bmod(n^2) \tag{11}$$

$$[t\cdot m\bmod n]=[m]^{t}\bmod(n^2) \tag{12}$$

1.3.2 室内定位中的Paillier算法的应用(PriWFL[16])

由式(4)可知,$d_{i,t}$求解可分解成$\sum_{j=1}^{M}f_{i,j}^2$、$\sum_{j=1}^{M}(-2f_{i,j}\cdot f_j)$和$\sum_{j=1}^{M}f_j^2$三个部分,其中第1部分$S_1=\sum_{j=1}^{M}f_{i,j}^2$是离线指纹的RSS信号,由服务器端完成相关操作;用户的定位信息包括S_2和S_3两部分,具体如式(13,14)所示,用户端将加密后的定位信息[S_2]和[S_3]上传至服务器。

$$S_2=\{-2f_1,-2f_2,\cdots,-2f_M\} \tag{13}$$

$$S_3=\sum_{i=1}^{M}f_i^2 \tag{14}$$

服务器根据用户上传的加密定位信息计算当前定位请求与每个参考点RP的加密欧氏距离,根据

式(11,12)可知,密文距离计算如式(15)所示。式中,$[d_{i,t}]$表示在线RSS信号与第i个RP的RSS信号的加密距离,R为服务器端选择的随机数,可在保持最近邻的同时,模糊RSS信号与各参考点的真实信号距离,有利于服务端的隐私保护。服务器计算加密距离后将其返回至用户端。用户端使用私钥对加密距离进行解密后,完成KNN定位。

$$\begin{cases}[d_{i,t}]=[S_{i,1}][S_{i,2}][S_{i,3}][R] \quad i\in[1,2,\cdots,N] \\ [S_{i,1}]=\left[\sum_{j=1}^{M}f_{i,j}^2\right] \\ [S_{i,2}]=\left[\sum_{j=1}^{M}(-2f_{i,j}\cdot r_j)\right]=\prod_{j=1}^{M}[(-2r_j)]^{f_{i,j}} \\ [S_{i,3}]=[S_3]\end{cases} \tag{15}$$

PriWFL算法解决了室内定位中的隐私保护问题,但由于Paillier算法为保证安全强度,其密钥长度一般为1 024 bit或更高,导致算法时间开销较大。根据式(13~15)可知,AP数量M影响用户端加密与服务器计算时间,RP数量N影响用户端解密与服务器计算时间,当参与定位的RP和AP数量增加时,获取定位服务的时间消耗较长,文献[18]中取AP数M为6,RP数N为200时,单次定位的总时间消耗达到10 s以上,这将降低定位的实时性,甚至不适用于移动用户的连续定位。

Paillier算法应用于室内定位中的信息传输框图如图1所示。

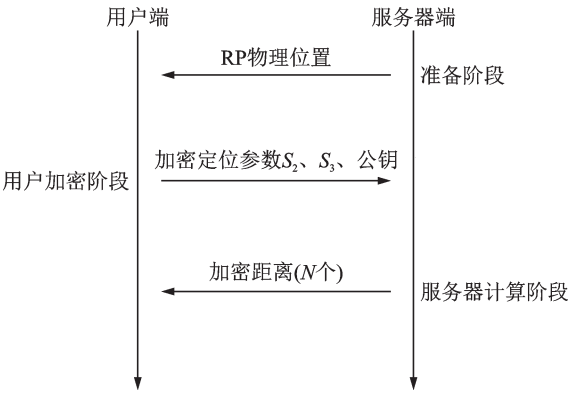


图1 室内定位中的Paillier算法的数据通信图

Fig.1 Data communication diagram for the Paillier algorithm in indoor positioning

2 系统模型、威胁模型及设计目标

2.1 系统模型

本文的系统模型如图2所示,由两部分实体组成:定位服务器与待定位用户。

(1) 定位服务器:负责生成室内地图并采集生成RSS指纹数据库,并将相关地图信息发送至待定位用户;负责计算用户指纹与每个RP指纹的密文域距离并返回。

(2) 待定位用户:负责生成密钥,并将公钥上传至服务器;采集RSS信号,经公钥加密后发送至服务器。

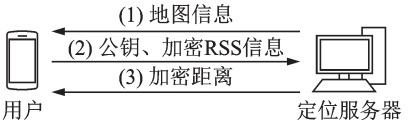


图2 系统模型

Fig.2 System model

2.2 威胁模型

本文设定用户与服务提供商都遵循半诚实,即诚实但好奇(Curious-but-honest)的安全模型[22],用户和服务提供商均诚实地遵守所设计的协议,但都打算公开对方的私人信息。本文着重于用户端的隐私保护即假设用户完全可信,服务器不可信。得出如下3种攻击威胁模型:

(1)用户请求攻击:服务器获取用户的RSS信号,通过RSS信号推测用户的位置。

(2)用户位置攻击:服务器直接从请求中获取用户的位置。

(3)明文信息攻击:服务器通过其他明文信息推测用户的位置。

2.3 设计目标

针对上述系统模型与威胁模型,设计目标如下:

(1) 隐私需求包括定位过程中服务器无法获知或推测用户所处位置;定位过程中服务器无法获知或推测用户运动轨迹。

(2) 定位精度:为满足定位需求,在保证用户隐私安全的前提下,本方案应达到或接近无隐私保护时定位算法的精度。

(3) 定位效率:在不降低定位精度的前提下,为适应实时定位的场景,本方案中的定位时间复杂度应尽可能低。

(4) 通信开销:本方案定位过程中用户与服务器间需进行通信,为保证定位效率,本方案中的通信开销应尽可能低。

3 移动对象室内定位中的隐私保护算法

针对加密定位时间消耗过长的问题,本文提出一种移动对象室内定位中的隐私保护算法,本算法中首先将定位分成连续位置定位与非连续位置定位。连续位置定位时利用前一次定位请求估计位置的信息减少候选RP与候选AP;同时这种连续位置定位过程能有效排除由于RSS信号突发大噪声时引起过大的定位误差。非连续位置定位时利用粗定位的方式减少候选RP与候选AP,进而提高定位效率。所提算法中,用户无需与服务器进行信息交换,利用自身采集的RSS信号即可完成两种定位方式的判定,从而获取自身位置附近的候选RP和候选AP信息,降低用户对自身定位请求指纹加密的计算开销,同时服务器端可以根据候选AP信息,对指纹库进行重构,来降低加密定位运算的计算开销。算法框架如图3所示,虚线框中为可选部分,当整个定位区域内的AP数较多时,利用PCA算法降维,在保留定位信息的前提下进一步减少计算的复杂度。

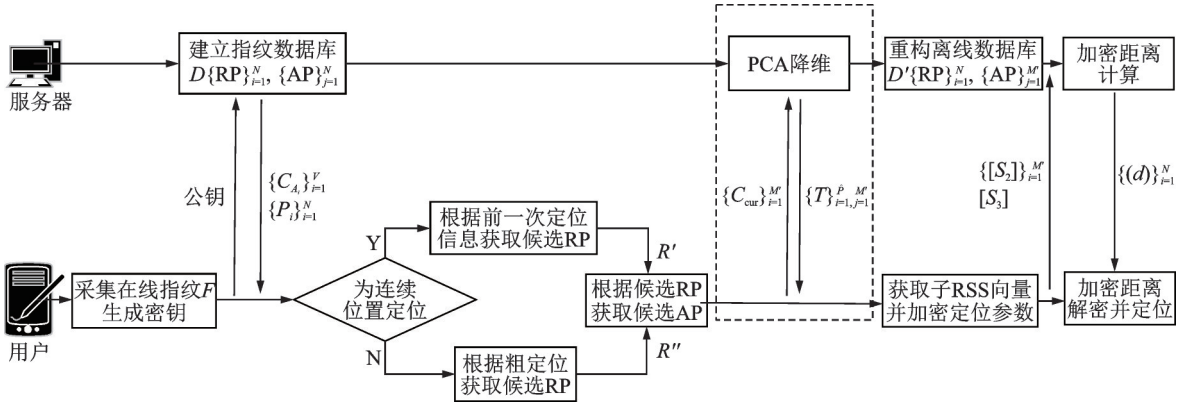


图3 本文算法整体框图

Fig.3 Overall block diagram of the proposed algorithm

3.1 服务器准备阶段

服务器将室内定位区域划分成大致均匀的V个子区域,表示为$\{A_i\}_{i=1}^{V}$,根据每个子区域内所有RP能接收到的AP确定子区域的覆盖向量C_{A_i}。在用户请求定位时,服务器将每个子区域的覆盖向量及每个子区域中包含的RP编号和位置发送给用户。整个定位时间内服务器准备阶段只需执行一次。

3.2 连续位置定位请求判定

在WiFi指纹定位中,通常利用的是RSS信号的强弱分布,而实验观察表明指纹库中AP信号的覆盖向量同样可提供定位相关信息。图4表示两个不同AP的信号热力图。从图4可观察到这两个AP的信号能到达的区域基本不重合,具有明显不同的覆盖区域。这也说明若两个位置的覆盖向量相似则其对应的物理位置相邻。

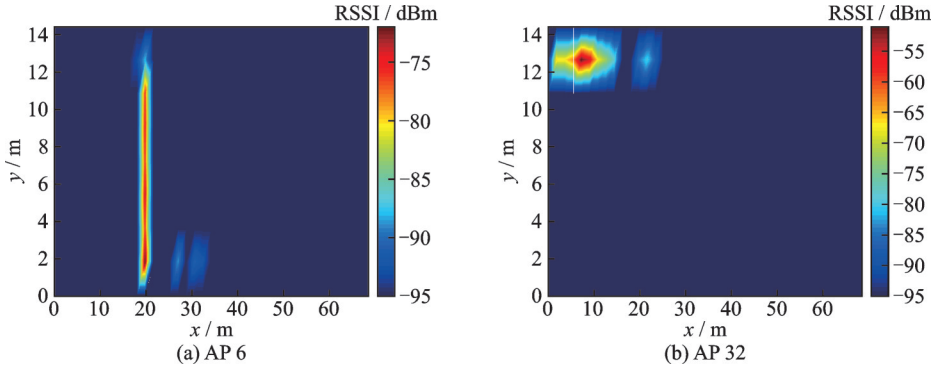


图4 指纹RSS信号热力图

Fig.4 Fingerprint RSS signal heat map

本文算法中用户通过计算当前位置的覆盖向量与前一次定位请求时的覆盖向量的相似度,也即汉明距离,判断当前定位请求是否为连续位置定位。假设当前位置的RSS信号为φ_t,对应的AP覆盖向量为$\boldsymbol{C}_t$,上一次定位请求的RSS信号为φ_{t-1},对应的AP覆盖向量表示为$\boldsymbol{C}_{t-1}$,则$\boldsymbol{C}_t$与$\boldsymbol{C}_{t-1}$之间的汉明距离可表示为

$$d_c(\boldsymbol{C}_t, \boldsymbol{C}_{t-1}) = \sum_{j=1}^{M} \left| c_{tj} - c_{(t-1)j} \right| \tag{16}$$

式中d_c表示汉明距离。此距离越小,表明当前定位请求与上一次定位请求能接收到的AP信号越一致,则在一般情况下其物理位置也较接近。当覆盖向量的差异小于某个阈值时,可认为是连续位置定位,即$d_c/M \leqslant \xi$时认为当前定位请求与上一次定位请求为连续位置定位,反之为非连续位置定位。此判定方法中用户只需利用自身采集信息、存储的信号,而无需与服务器进行信息传送。

3.3 移动对象位置定位

如图3所示,移动对象位置定位分5个阶段:获取候选RP、重构RSS向量、用户端加密、服务器端计算和用户端解密并定位。

(1) 获取候选RP集合

①连续位置定位:当前定位为连续位置定位时则用户当前实际位置接近于上次定位请求的实际位置。假设上次定位请求的估计位置较准确,取距上一次定位请求的估计位置$\hat{P}_{\text{pre}}$最近的X个RP并随机选取Y个匿名RP作为当前定位请求的候选RP。匿名RP可干扰服务器对用户位置的判定,降低用户端的位置隐私泄露。取$N'=X+Y$,则连续位置定位的候选RP集合可表示为$R'=\{\text{RP}'_1, \text{RP}'_2, \cdots, \text{RP}'_{N'}\}$。

②非连续位置定位:如式(16)所示,计算当前定位请求的AP覆盖向量与每个子区域的AP覆盖向量的汉明距离,取出汉明距离最小的a个子区域编号,即用户粗定位在这a个子区域中。为进一步降低

隐私的泄漏，在这a个子区域外任意选取b个匿名子区域，匿名子区域可干扰服务器对用户位置的判定，降低用户端的位置隐私泄露。假设$(a+b)$个子区域的编号对应的子区域集合为$A'=\{A'_1, A'_2, \cdots, A'_{(a+b)}\}$，子区域集合中的RP也即候选RP个数为$N''$，集合表示为$R''=\{\mathrm{RP}''_1, \mathrm{RP}''_2, \cdots, \mathrm{RP}''_{N''}\}$。

(2) 重构RSS向量

取集合R(R'或R'')中候选RP能采集到的AP并集作为参与当前定位请求的候选AP，如式(17)所示，若$C'_{\mathrm{cur},i}=1$表示第i个AP参与当前定位，$C'_{\mathrm{cur},i}=0$表示第i个AP不参与当前定位，候选AP中包含当前用户能采集到RSS信号的所有AP。

$$C'_{\mathrm{cur}}=C_{\mathrm{RP}'_1}\cup C_{\mathrm{RP}'_2}\cup\cdots\cup C_{\mathrm{RP}'_{N_e}} \tag{17}$$

式中N_e为N'或N''。假设候选AP总数为M'，取出φ_t中$C'_{\mathrm{cur},i}=1$的部分构成当前定位请求的在线RSS子指纹$\boldsymbol{\varphi}'_t$，具体可表示为

$$\boldsymbol{\varphi}'_t=[f'_1, f'_2, \cdots, f'_{M'}] \tag{18}$$

(3) 用户端加密

计算得到当前定位请求的RSS子信号$\boldsymbol{\varphi}'_t$后，对定位参数进行加密得到$[S'_2]$，$[S'_3]$后与公钥、C'_{cur}一起上传至服务器。

(4) 服务器端计算

服务器根据用户上传的C'_{cur}，重构当前定位请求的子数据库，重构规则为取出数据库中对应$C'_{\mathrm{cur},i}=1$的部分AP，也即候选AP为

$$\boldsymbol{\varphi}_i{}'=[P_i', F_i']=[x'_i, y'_i, f'_{i,1}, f'_{i,2}, \cdots, f'_{i,M'}] \quad i\in[1, 2, \cdots, N] \tag{19}$$

服务器在密文域计算$\boldsymbol{\varphi}'_t$与$\boldsymbol{\varphi}'_i$的欧式距离$[d'_{i,t}]$，并将加密后的欧氏距离返回至用户端。通过指纹库的重构，服务器需要计算的指纹长度从M降低至M'，由此降低服务器端的计算开销。

(5) 用户端解密并定位

用户端接收到N个加密距离只需解密候选RP对应编号的加密距离，取距离最小的K个RP利用式(5)完成位置估计。根据上述描述可知，连续位置定位与非连续位置定位的用户端与服务器端的数据通信分别如图5、6所示。

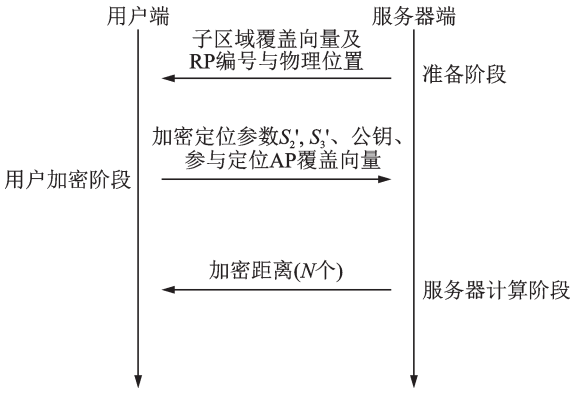


图5 连续定位用户端与服务器端的数据通信

Fig.5 Continuous positioning of user-side and server-side data communication

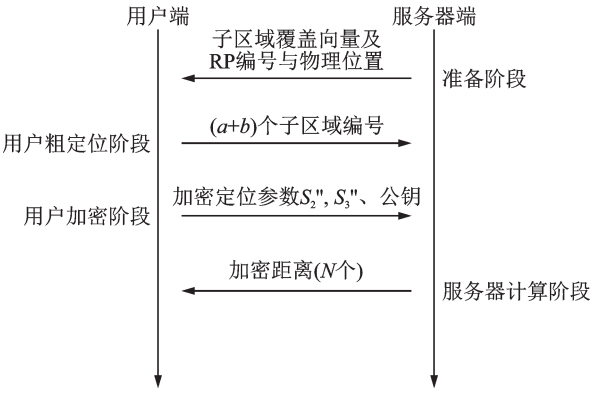


图6 非连续定位用户端与服务器端的数据通信

Fig.6 Discontinuous positioning of user-side and server-side data communication

3.4 PCA降维

密文域距离计算时,服务器根据用户上传的当前定位候选AP的覆盖向量C'_{cur},重构离线数据库。当实验环境中存在大量AP,导致用户所选的候选AP仍较多时,后续密文域最近邻检索依旧存在较大的计算开销,此时可利用PCA降维进一步提高定位效率。利用式(6)对重构后的子离线数据库进行降维得到$F_{i,P}$与转换矩阵T,并将转换矩阵T发送至用户端,用户端根据转换矩阵T对φ'_t降维后再进行加密运算,从而降低后续计算的复杂度,此过程并不涉及用户端或服务器端额外的隐私泄露。

4 安全性分析

针对2.2节中提出的威胁模型对本文所提算法进行安全性分析。

(1) 用户请求攻击。定位过程中用户需上传加密后的RSS指纹[S_2]和[S_3]至服务器。本文选用语义安全的Paillier密码系统[19],在服务器或恶意攻击者未获取私钥的情况下,无法对其进行解密而获得RSS指纹的明文信息。因此本文所提算法能抵抗用户请求攻击,用户的请求信息是安全的。

(2) 用户位置攻击:服务器通过密文计算可得到用户的RSS指纹与每个RP的RSS指纹之间的加密距离[d]。在服务器或恶意攻击者未掌握用户私钥的情况下,无法对其进行解密获得用户的近邻参考点。因此本文所提算法能抵抗用户位置攻击,用户的位置信息是安全的。

(3) 明文信息攻击:所提算法定位过程中用户需上传的明文信息是候选AP集的覆盖向量C'_{cur}。服务器可能通过其估计用户所在的局部区域。因此在所提算法中引入了候选RP集。该候选RP集所加入的匿名RPs可干扰服务器对用户位置的追踪。在本文5.4节中实验证明了服务器无法通过C'_{cur}估计用户所处的局部区域,因此也不可能进一步确定用户的位置。用户的位置信息是安全的。

5 实验结果及分析

为验证本文所提算法在实际室内定位中的定位性能,在大楼室内搭建约70 m×14 m的实验环境进行实际的定位测验,实验环境如图7所示,图7中圆点表示参考点,共92个,三角形表示在线测试点共42个,在整个实验环境中共能采集到199个AP。目标未能接收到信号的AP的RSS值设为0。红框表示子区域共16个,每个子区域中含有6~7个参考点。蓝线表示连续位置定位时用户的移动轨迹,连续两个在线测试点的距离约为1.8 m。非连续位置定位时,在线测试点的顺序随机,且连续两个在线测试点间的距离大于3 m。

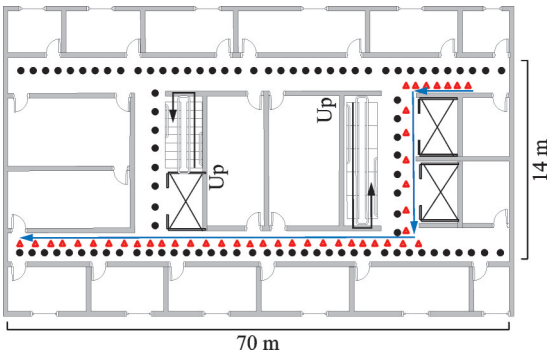


图7 实验布局图

Fig.7 Experimental layout diagram

5.1 参数选择对性能的影响

本文所提算法不需要进行额外的设备部署与模型训练,直接根据接在线接收到的RSS信号进行目标定位,但也需确定若干参数。实验中分析了如何选择合适的参数保证定位性能。

(1) 判断是否为连续定位的阈值ξ

阈值ξ直接决定了当前定位请求是否连续,由图7可知两条指纹的AP覆盖向量的汉明距离随着目标间的距离增加而增大,在本文中假设物理距离在2 m之内认为目标为连续定位,由图8可知当物理距离为2 m时,汉明距离为30左右,则取阈值ξ为0.15。

(2) 连续位置定位中的最近邻估计数X

在连续位置定位中最近邻估计数X实际决定了服务器进行最近邻所搜的范围,其值直接影响隐私保护的定位性能。如图9所示,实验中以第18个测试点为例,X值较小时由于并未覆盖最近邻,导致定位误差较大,当X增大至10时,此时已经覆盖当前定位请求的K个最近邻,因此可达到与KNN算法一致的定位精度,后续随着X的增加,定位误差基本无影响。但候选AP数随X的增加而增加,从而使得用户和服务器端计算开销增大。且长时间的连续位置定位,必然会带来一个误差累计的过程即上次定位请求的精度会直接影响当前定位请求的定位精度,因此需选择合适的X以减少误差累计,并降低计算开销。

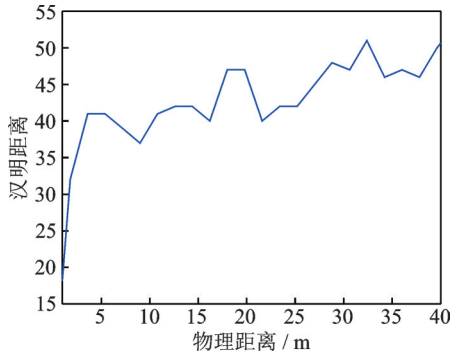
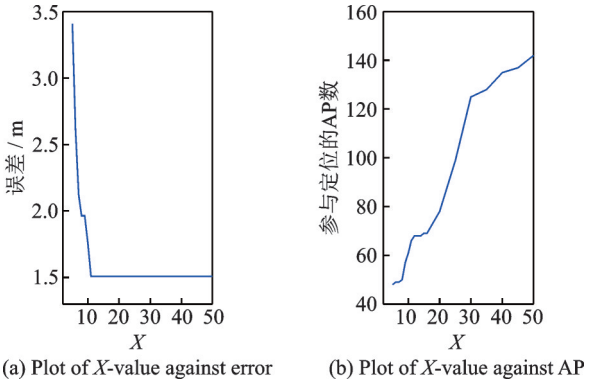


图8 覆盖向量汉明距离与物理距离的关系

Fig.8 Hamming distance of covering vector versus physical distance



(a) Plot of X-value against error (b) Plot of X-value against AP

图9 X值与定位性能关系图

Fig.9 Relationship between X-value and positioning performance

如图10所示,在实验中当X小于10时,连续位置定位带来的误差累计随着定位请求次数的增加而急剧上升至40 m以上,当X大于10时,连续位置定位带来的误差累计可以忽略不计,并且在第25、第34次定位请求时,原算法出现了大误差的现象,但是在本文中所提出的方法中取X为10时,解决了大误差的现象。这是由于X的存在,当前定位请求与上一次定位请求产生联系,即使在某次信号采集时出现干扰也可利用上一次定位的信息进行修正,如图11所示为连续位置定位的定位误差的累积分布函数(Cumulative distribution function, CDF)图,当取X=10时,实验中93%的测试点定位误差小于4 m,并且消除了大于6 m的定位误差,而在KNN算法中只有85%的测试点定位误差小于4 m,最大定位误差达7 m以上。

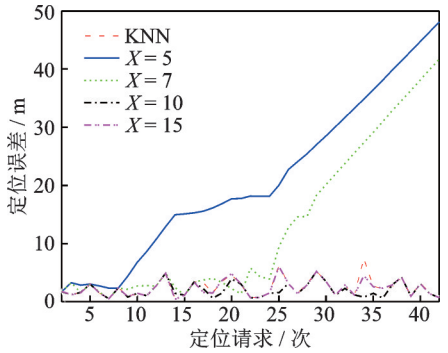


图10 连续位置定位的误差与X值关系图

Fig.10 Error versus X-value for continuous position positioning

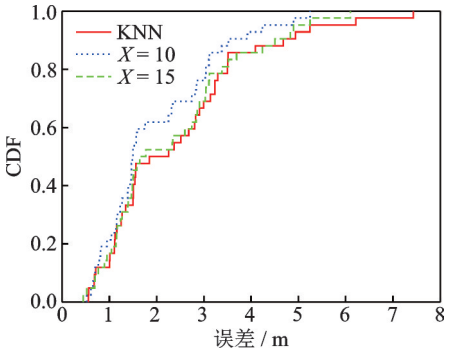


图11 连续位置定位的定位误差的CDF图

Fig.11 CDF diagram of the positioning error for continuous position positioning

(3) 非连续位置定位中子区域数 a

在非连续位置定位中取 a 个子区域与 b 个匿名子区域内的RP作为当前定位请求的候选RP,取候选RP能接收到RSS信号的AP并集作为当前定位请求的候选AP。如图12可知,取 a=3时,定位精度与KNN算法相当。

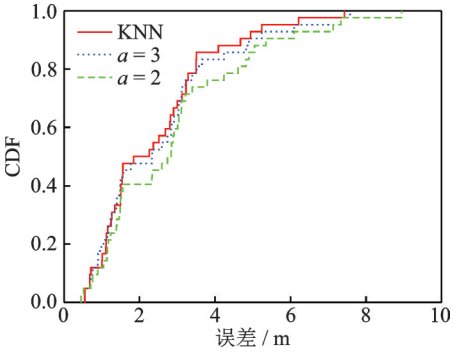


图12 非连续位置定位的定位误差的CDF图

Fig.12 CDF diagram of positioning error for discontinuous position positioning

5.2 与PriWFL方法的比较

5.2.1 定位性能比较

连续位置定位中取 X=10、Y=3时的定位性能与PriWFL算法的对比如表1所示,其中U表示用户端,S表示服务器端;非连续位置定位中取 a=3,b=3时的定位性能与PriWFL算法性能对比如表2所示。由表1、2可知,本文所提出的算法不管是连续定位或是非连续定位均能保证定位误差基本不变或定位误差更好的情况下实现对定位效率的大幅度提升,连续位置定位中用户端效率提高75%,单次定位所需时间为1 s,服务器端提高65%,并且消除了大误差的影响。整体定位精度提升20%,非连续定位中用户端效率提高40%,单次定位所需时间为0.52 s,服务器端定位效率提高60%,定位误差基本不变。添加PCA算法后,定位精度基本无影响,定位效率大幅提升。

表1 连续位置室内定位算法与PriWFL算法性能对比(X=10,Y=3)

Table 1 Performance comparison of continuous indoor positioning algorithm and PriWFL algorithm (X=10, Y=3)

算法	定位误差/m	加密时间(U)/s	计算时间(S)/s	解密时间(U)/s	单次定位请求总时间/s
PriWFL	2.44	1.13	12.02	0.51	13.95
无PCA	2.42	0.81	5.6	0.19	6.6
PCA	2.45	0.18	2.3	0.19	2.67

表2 非连续室内定位算法与PriWFL算法性能对比(a=3、b=3)

Table 2 Performance comparison of discontinuous indoor positioning algorithm and PriWFL algorithm (a=3, b=3)

算法	定位误差/m	加密时间(U)/s	计算时间(S)/s	解密时间(U)/s	单次定位请求总时间/s
PriWFL	2.44	1.13	12.02	0.51	13.95
无PCA	2.03	0.46	5.6	0.06	6.02
PCA	2.10	0.18	2.3	0.06	2.54

5.2.2 通信开销比较

在本文实验中,信号接入点AP数 M 为199,参考点RP数 N 为92,实验环境中子区域数 V 为16,明文数据加密后的密文数据大小为1 024 bit,连续位置定位中的候选RP数 X 为10,匿名RP数 Y 为3,非连续位置定位中候选子区域 a 为3,匿名子区域 b 为3,PCA降维后的主成分个数 $\hat{p}$ 为30。以第2个定位点为例计算通信开销,非连续位置定位的候选AP数 M' 为90,连续位置定位的候选AP数为50。服务器与用户间具体通信信息与通信开销对比如表3所示,表中S表示服务器向用户发送信息、U表示用户向服务器上传信息。

综上,本文所提算法中无PCA降维时,连续位置定位与非连续位置定位的通信开销较PriWFL算法分别减少了50%和40%。添加PCA降维后,通信开销减少了55%。

表3 本文算法与PriWFL算法的通信开销对比

Table 3 Comparison of communication overhead between the proposed algorithm and the PriWFL algorithm

bit

算法	通信信息	理论通信开销	实验通信开销
PriWFL	RP物理位置(S) 公钥及加密定位参数$[S_2][S_3]$(U) 加密距离$[d]$(S)	$2\times N\times 1$ $(M+1)\times 1\,024+1$ $N\times 1\,024$	$293\times 1\,024$
无PCA	子区域覆盖向量、RP物理位置及编号(S) 公钥、候选AP集合$\boldsymbol{C}_{\text{cur}}$、加密定位参数$[S_2'][S_3']$(U) 加密距离$[d]$(S)	$(M\times V)+(2\times N)$ $M'\times(1\,024+1)+1\times 1\,024$ $N\times 1\,024$	非连续 $175\times 1\,024$ 连续 $145\times 1\,024$
PCA	子区域覆盖向量、RP物理位置及编号(S) 公钥、候选AP集合$\boldsymbol{C}_{\text{cur}}$(U) 转换矩阵$\boldsymbol{T}$(S) 加密定位参数(U) 加密距离$[d]$(S)	$(M\times V)+(2\times N)\times 1$ $(M'+1)\times 1$ $M'\times\hat{p}\times 1$ $\hat{p}\times 1\,024$ $N\times 1\,024$	$135\times 1\,024$

5.3 公共数据集验证

本文还利用公共数据集UJIIndoorLoc[23]中1F的数据对实验结果进行验证。公共数据集中数据经处理后,在线测试点共90个,参考点RP共307个,整个实验环境共32个AP,并将整个定位区域均分成20个子区域。与实际测验数据集类似,使用公共数据集对非连续位置定位与连续位置定位两种定位方式进行定位结果验证分析。验证实验中取最近邻估计数X为50,匿名RP数Y为15,子区域数a为4,匿名区域数b为4,绘制定位误差的CDF图,分别如图13、14所示。由图13,14可知,公共数据集中非连续位置定位的定位精度基本与KNN算法一致,而连续位置定位中的定位精度略优于KNN算法,并且解决了定位过程中产生的大误差现象。综上,本文所提算法在公共数据集与实际测验数据集中的实验结果一致,且由文献[19]中实验结果可知,文献[19]所提算法定位精度低于KNN算法,因此可知本文所提算法定位效果优于文献[19]中所提算法。

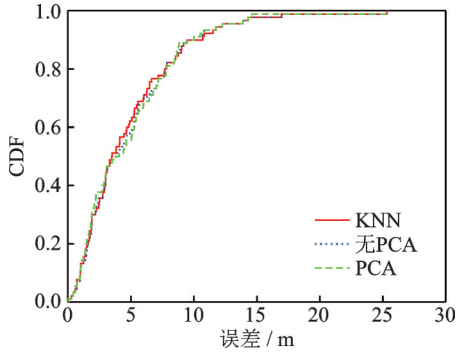


图13 公共数据集非连续位置定位的定位误差的CDF图

Fig.13 CDF diagram of positioning error for discontinuous position positioning of public datasets

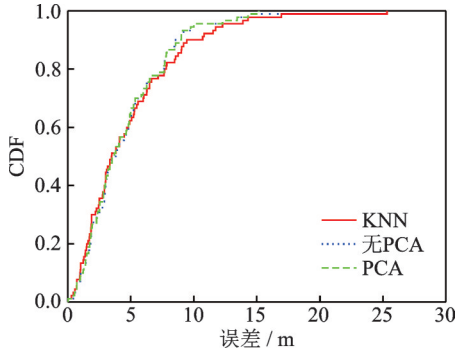


图14 公共数据集连续位置定位的定位误差的CDF图

Fig.14 CDF diagram of positioning error for continuous position positioning of public datasets

5.4 隐私保护分析

为证明服务器无法通过候选AP集估计用户所处的位置。以实验数据库中连续位置定位中的第2个定位点为例,取X为10,匿名RP数Y为3。由于服务器从用户端获取的明文信息仅为,即用于定位的AP列表,包含了用户真实位置的最近邻RP和匿名RP所能获取的所有AP。假设当某一RP能接收到70%的候选AP时,服务提供商认为该RP在用户可能所在的子区域内,实验结果如图15所示,绝大部分RP被服务提供商判定在用户可能所在的子区域内,模糊了用户的真实位置。可见,服务器仅根据候选AP集合并无法获取用户当前所处的位置或子区域。用户在定位区域内连续定位时,由于服务器无法获取用户单次定位的位置或子区域,并且相邻的两次定位中并无其他的定位信息泄露,因此服务器无法获取到移动对象的移动轨迹,实现了对移动对象的轨迹保护。而在非连续位置定位中,服务器能获取到的用户定位信息与连续位置定位类似,即只能获取候选AP集,因此,非连续位置定位时,服务器也不能获取用户当前所处的位置或子区域。保护了用户的位置隐私。在公共数据集中也得到了同样的实验结果,即服务器无法通过候选AP集获取用户当前所处的位置或子区域。

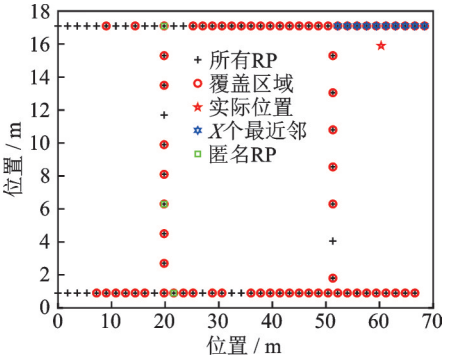


图15 候选AP的RSS信号覆盖区域图

Fig.15 RSS signal coverage area map of candidate APs

6 结束语

针对Paillier算法应用于室内定位实现隐私保护时定位效率低的问题,提出一种移动对象在室内定位中的隐私保护算法。用户端首先判断当前定位请求是否为连续位置定位,若为连续位置定位则通过上一次定位请求估计位置信息减少候选RP数,非连续位置定位则通过粗定位方式减少候选RP数;其次利用候选RP选取候选AP并重构在线RSS信号向量与离线数据库,达到减少参与定位运算的RP数与AP数的目的,并提供一种可选的PCA降维方案。最终算法实现了保证定位精度与用户隐私,且在不增加通信开销的前提下,提高室内定位的效率并实现移动对象的轨迹匿名。

参考文献:

[1] LABINGHISA B, PARK G S, LEE D M. Improved indoor localization system based on virtual access points in a Wi-Fi environment by filtering schemes[C]//Proceedings of 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN). Sapporo, Japan: IEEE, 2017.

[2] LI Shenghong, HEDLEY M, BENGSTON K, et al. Passive localization of standard WiFi devices[J]. IEEE Systems Journal, 2019, 13(4): 3929-3932.

[3] YU N, ZHAN X H, ZHAO S N, et al. A precise dead reckoning algorithm based on bluetooth and multiple sensors[J]. IEEE Internet of Things Journal, 2018, 5(1): 336-351.

[4] 孙伟,李亚丹,黄恒,等. 基于级联滤波的建筑结构信息/惯导室内定位方法[J]. 仪器仪表学报, 2021, 42: 10-16.
SUN Wei, LI Yadan, HUANG Heng, et al. Building structure information/inertial navigation indoor positioning method based on cascade filtering[J]. Chinese Journal of Scientific Instrument, 2021, 42: 10-16.

[5] 米伟娟,李娜. 基于变分自编码器的WLAN定位方法[J]. 电子测量与仪器学报, 2020, 34(12): 101-108.
MI Weijuan, LI Na, WLAN localization method based on variational autoencoder[J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(12): 101-108.

[6] MAHFOUZ S, MOURAD-CHEHADE F, HONEINE P, et al. Non-parametric and semi-parametric RSSI/distance

modeling for target tracking in wireless sensor networks[J]. IEEE Sensors Journal, 2016, 16(7): 2115-2126.

[7] BAHL P, VENKATA N, RADAR P. An in-building RF-based user location and tracking system[C]//Proceedings of Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Piscataway, USA: IEEE, 2002: 775-784.

[8] JIANG H B, ZHAO P, WANG C. RobLoP: Towards robust privacy preserving against location dependent attacks in continuous lbs queries[J].IEEE/ACM Transactions on Networking, 2018, 26(2): 1018-1032.

[9] CHINTALAPUDI K, IYER A P, PADMANABHAN V N. Indoor localization without the pain[C]//Proceedings of ACM/IEEE International Conference on Mobile Computing and Networking. [S.l.]: ACM, 2010.

[10] ZHAO P, LI J, ZENG F Z, et al. ILLIA: Enabling k-anonymity-based privacy preserving against location injection attacks in continuous LBS queries[J]. IEEE Internet of Things Journal, 2018, 5(2): 1033-1042.

[11] ZHANG Y, TONG W, ZHONG S. On designing satisfaction-ratio-aware truthful incentive mechanisms for k-anonymity location privacy[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(11): 2528-2541.

[12] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//Proceedings of the First International Conference on Mobile Systems, Applications, and Services (MobiSys 2003). San Francisco, CA, USA:[s.n.], 2003.

[13] BERESFORD A R, STAJANO F. Mix zones: User privacy in location-aware services[C]//Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops. [S.l.]: IEEE, 2004: 215-233.

[14] PAILLIER P. Public-key cryptosystems based on composite degree residuosity classes[J]. Lecture Notes in Computer Science, 1999, 1592(1): 223-238.

[15] HOFHEINZ D, JAGER T. Tightly secure signatures and public-key encryption[J]. Designs Codes and Cryptography, 2016, 80(1): 29-61.

[16] LI H, SUN L, ZHU H, et al. Achieving privacy preservation in WiFi fingerprint-based localization[C]//Proceedings of IEEE INFOCOM IEEE Conference on Computer Communications. [S.l.]: IEEE, 2014.

[17] ZHANG T, CHOW SHERMAN S M, ZHOU Z, et al. Privacy-preserving Wi-Fi fingerprinting indoor localization[J].Lecture Notes in Computer Science, 2016, 9836: 215-233.

[18] HIGUCHI T, MARTIN P, CHAKRABORTY S, et al. AnonyCast: Privacy-preserving location distribution for anonymous crowd tracking systems[C]//Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UBICOMP 2015). [S.l.]: ACM, 2015: 1119-1130.

[19] ZHANG Guanglin, ZHANG Anqi, ZHAO Ping, et al. Lightweight privacy-preserving scheme in Wi-Fi fingerprint-based indoor localization[J]. IEEE Systems Journal, 2020, 14(3): 4638-4647.

[20] FANG S H, LIN T N.Principal component localization in indoor wlan environments(Article)[J]. IEEE Transactions on Mobile Computing, 2012, 11(1): 100-110.

[21] 乐燕芬,汤卓,盛存宝,等. 基于多分布密度位置指纹的高效室内定位算法研究[J].通信学报, 2019, 40(1): 172-179.
LE Yanfen, TANG Zhuo, SHENG Cunbao, et al. Research on efficient indoor positioning algorithm based on multi-distribution density location fingerprint[J]. Journal on Communications, 2019, 40(1): 172-179.

[22] ROY P, CHOWDHURY C G, BANDYOPADHYAY S. JUIndoorLoc: Aubiquitous framework for smartphone-based indoor localization subject to context and device heterogeneity[J]. Wireless Personal Communications, 2019, 106(2): 739-762.

[23] YANG, Q, LIU Y, CHEN T J, et al. Federated machine learning: Concept and applications[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 12.

作者简介:

张志武(1999-),男,硕士研究生,研究方向:室内定位与隐私保护,E-mail:1547231823@qq.com。

雷若兰(1999-),女,硕士研究生,研究方向:室内定位、信息隐藏,E-mail:1665844775@qq.com。

乐燕芬(1978-),**通信作者**,女,博士,副教授,研究方向:室内定位、无线网络和多媒体,E-mail:leyanfen@usst.edu.cn。

(编辑:陈珺)